Spatiotemporal monthly rainfall reconstruction via artificial neural network ? case study: south of Brazil

P. S. Lucio, F. C. Conde, I. F. A. Cavalcanti, A. I. Serrano, A. M. Ramos, A. O. Cardoso

Adv. Geosci., 10, 67–76, 2007
www.adv-geosci.net/10/67/2007/

Advances in Geosciences

Spatiotemporal monthly rainfall reconstruction via artificial neural network – case study: south of Brazil

P. S. Lucio[1,2], F. C. Conde[1,3], I. F. A. Cavalcanti[4], A. I. Serrano[1], A. M. Ramos[1], and A. O. Cardoso[4]

[1]Centro de Geofísica de Évora (CGE), Universidade de Évora, Apartado 94, 7002-554 Évora, Portugal
[2]Departamento de Estatística (DEST), Universidade Federal do Rio Grande do Norte (UFRN), Brazil
[3]Instituto Nacional de Meteorologia (INMET), Brasília DF, Brazil
[4]Centro de Previsão do Tempo e Estudos Climáticos (CPTEC/INPE), Cachoeira Paulista SP, Brazil

Received: 14 September 2006 – Revised: 5 January 2007 – Accepted: 31 January 2007 – Published: 26 April 2007

Abstract. Climatological records users, frequently, request time series for geographical locations where there is no observed meteorological attributes. Climatological conditions of the areas or points of interest have to be calculated interpolating observations in the time of neighboring stations and climate proxy. The aim of the present work is the application of reliable and robust procedures for monthly reconstruction of precipitation time series. Time series is a special case of symbolic regression and we can use Artificial Neural Network (ANN) to explore the spatiotemporal dependence of meteorological attributes. The ANN seems to be an important tool for the propagation of the related weather information to provide practical solution of uncertainties associated with interpolation, capturing the spatiotemporal structure of the data. In practice, one determines the embedding dimension of the time series attractor (delay time that determine how data are processed) and uses these numbers to define the network's architecture. Meteorological attributes can be accurately predicted by the ANN model architecture: designing, training, validation and testing; the best generalization of new data is obtained when the mapping represents the systematic aspects of the data, rather capturing the specific details of the particular training set. As illustration one takes monthly total rainfall series recorded in the period 1961–2005 in the Rio Grande do Sul – Brazil. This reliable and robust reconstruction method has good performance and in particular, they were able to capture the intrinsic dynamic of atmospheric activities. The regional rainfall has been related to high-frequency atmospheric phenomena, such as El Niño and La Niña events, and low frequency phenomena, such as the Pacific Decadal Oscillation.

Correspondence to: P. S. Lucio
(pslucio@uevora.pt, pslucio@ccet.ufrn.br)

1 Introduction

One of the major problems in examining weather records for detecting changes in extremes is the lack of high-quality, long-term data (ground-based meteorological network does not operate over a common time period of adequate length). In general, the biggest drawback is that recorded data available must be gap-filled and quality controlled to provide a reliable continuous reference time series. It is important for time periods where no satisfactory reference series can be built due to insufficient number of suitable nearby stations or large discontinuities in the time series. Good quality database undoubtedly provides a key source of historical meteorological information for detection and monitoring of climate variability. However, in general, the meteorological network was not designed to serve this function, and preliminary evaluations indicate that few weather stations meet the criteria necessary for inclusion in a climatological sub-network.

The spatial distribution precipitation is summarised by the subjective descriptive four-moment measures: Mean, Variance, Skewness and Kurtosis, giving support to the spatial pattern recognition. A number of homogeneity tests (testing for structural stability) to detect non homogeneities were employed. Methods for all blended and the effect of natural variability is established taking into account ensembles of consecutive years currently used are: (1) The Pettitt test by A. N. Pettitt (1979) also called Mann-Whitney-Pettitt method and the Chow test by G. C. Chow (1960) were selected for nonparametric approach based on the Mann-Whitney; (2) The SNHT – developed by Alexandersson (1986, 1995) and Alexandersson and Moberg (1997); (3) The Range Buishand test by T. Buishand (1982); (4) The Von Neumann ratio test by J. von Neumann (1941) and (5) The Craddock test by J. M. Craddock (1979)).

Artificial Neural Network (ANN) procedures are increasingly used in climatological applications (Kalogirou et al., 1997; Michaelides et al., 1995; Abdelaal and Elhadidy, 1995;

Published by Copernicus GmbH on behalf of the European Geosciences Union.

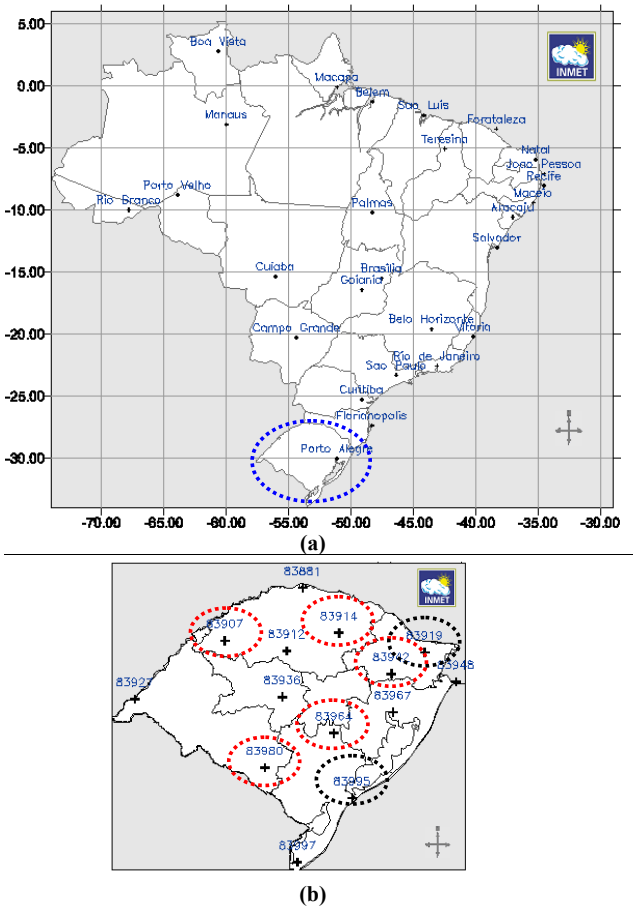

Fig. 1. **(a)** Brazilian political map and the target region (ellipse marked in blue). **(b)** Target and control rainfall stations used in this study. Positions of stations are indicated by cross, WMO numbered as referred to in the text. The meteorological stations of control (ellipses marked in red) and meteorological stations with larger number of "missing values" (ellipses marked in black).

Schizas et al., 1994). The ANN can provide proper solutions for climatological problems that are characterised by non-linearities. The idea in using neural networks or any other stochastic method for estimating missing rainfall values is considered because the rainfall recorded at any particular period at the target and its respective control stations determines a state in the space-time domain that can be emulated from past or future states. (cf., Kalogirou et al., 1997; Michaelides et al., 1995)

As expected, this robust reconstruction method has a good performance; since more information is introduced in the decision-making system (the conclusion highlights the use of climate proxies response as potential weather predictor). We capture the intrinsic dynamics of atmospheric activities, reproducing good long-term forecasting for periods of at least a complete cycle of the "El Niño South Oscillation" (ENSO), the Pacific Decadal Oscillation (PDO) and the Pacific/North American Teleconnection Pattern (PNA). It seems that the dynamics are essentially non-chaotic in this time scale, but perturbed by a fairly large amount of noise. Moreover, some meteorological variables over Brazil could be accurately predicted taking into account the model developed by artificial neural network. This approach recognises very well the mutual spatiotemporal rainfall variability dependence. In addition, the knowledge of phenomena connected to the precipitation variability is very important, particularly where the cases of extreme precipitation events affect negatively the life of the populations provoking flooding and dislodgement of families, or droughts that deprive them of essentials resources of subsistence.

The purpose of this work is to obtain homogeneous climatological series starting from a set of meteorological attributes. A reconstruction criteria based on the construction of an artificial neural network was adopted. This procedure was used to substantiate the obtained values integrated in the spatiotemporal structure of the time series, making possible to calculate the error estimated by each predicted value of the new series. The technique can be used to fill-in missing data from the rainfall observation network but also for checking suspected data by using the records from surrounding stations. This work also consists in analysing long-term observed rainfall series for localities of some Brazilian regions, corroborating the spatial consistency, apparent cycles and respective trends. The regional rainfall has been related to interannual variability, such as El Niño and La Niña events (Kousky et al., 1984; Grimm et al., 2000); and low frequency phenomena, such as the PDO (Andreolli and Kayano, 2005) A complementary application was carried out correlating the "monitored" rainfall database with the National Center for Environmental Prediction (NCEP) – National Center for Atmospheric Research (NCAR) reanalysis dataset to verify possible divergences relative to the observed data.

In a first phase, with the objective of illustrating the method, 6 meteorological stations (83907, 83914, 83950, 83967, 83919 and 83995), of the State of the Rio Grande do Sul – Brazil were selected (Fig. 1). They were considered representative of the climate variability of the area, esteeming and filling out the "missing values" in the series of accumulated monthly precipitation. It is important to emphasise that ANN models can be trained to determine the best mathematical relationship between the atmospheric circulation and the regional climate, without pre-defined restrictions. This methodology is capable to capture some of the nonlinear relationships between the local climate and the atmospheric circulation in large scale. The utility of ANN models lies in the fact that they can be used to infer a function from observations.

2 Time series reconstruction background

Data reconstruction is a methodology developed by climate scientists and meteorologists to remove inconsistencies in a

time series due to factors unrelated to weather, such as station location change, station environment change or change in instrumentation. The objective of this work is to access the availability, reliability and homogeneity of the historical series of meteorological data. The developments of a continuous and complete monthy dataset are useful in a variety of meteorological and hydrological research applications.

In Eischeid et al. (2000) six different methods of spatial interpolation were used to create a complete serial dataset for the western United States (all states west of the Mississippi River). It includes 2034 minimum-maximum temperature stations and 2962 total daily precipitation locations. The methods were: (1) the normal ratio method (NR); (2) simple inverse distance weighting (IDW); (3) optimal interpolation (OI); (4) multiple regression using the least absolute deviation criterion (MLAD); (5) the single best estimator; and (6) the median (MED) of the previous five methods (Eischeid et al., 1995). The interpolation schemes were evaluated by monthly integration method. The cross-validation of the results indicated a distinct seasonality to the efficiency of the estimates, although no systematic bias in the estimation procedures was found. Statistical summaries were generated using cross correlations between observed daily values and those estimated for each of the six different methods described. The six techniques respond to variations in season and geography, and the best estimation method is selected based on the efficiency of the estimate over time. The cross correlations were used to measure the efficiency of each method, and the method that exhibits the highest correlation relative to the other methods is utilized to replace missing values.

Additional investigations performed by the Northeast Regional Climate Center (DeGaetano et al., 1993) have shown that regression based methods of data estimation tend to be more accurate than within-station methods. An additional work (Huth and Nemesova, 1995) has shown that other weather elements, such as relative humidity, wind speed, and cloudiness, contribute very little to regression-based methods and that temperature at neighbouring stations has by far the highest spatial correlations. DeGaetano et al. (1993) go on to mention that "while such methods are useful over limited areas, they are computationally intensive and therefore not feasible when data estimates are needed for a large number of stations over a long period of time". These limitations have been partially overcome with the use of new high-speed workstations and large mass storage capabilities that now provide the horsepower required to perform these intensive calculations in a reasonable time period. In effect, Statistics is changing. Modern computers and software make it possible to look at data graphically and numerically in ways previously inconceivable. The Artificial Neural Network methods are part of this revolution.

3 Artificial Neural Network (ANN)

Time series is a special case of symbolic regression and can be done using the framework of mathematical modelling by an artificial intelligence network (Bishop, 1995). The Artificial Neural Network (ANN) explores the dependence of meteorological attributes as a function of space and time on inputs to the computer simulations. The use of ANN has been recognized recently as a promising way of making estimations on time series, detecting irregular behaviour. Because estimates are required for each sample unity separately over a variety of terrain with a differing number of available surrounding observations, we have chosen a different method for filling meteorological gaps. An ANN is an interconnected group of artificial neurones that uses a mathematical model for information processing based on a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network. In more practical terms neural networks are non-linear statistical data modelling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data.

The ANNs are essentially simple mathematical models defining a function $f:X\rightarrow Y$. Each type of ANN model corresponds to a class of such functions. In effect, the word network in the term "artificial neural network" arises because the function $f(X)$ is defined as a composition of functions $g_k(X)$, which can further be defined as a composition of other functions. This can be conveniently represented as a network structure, with arrows depicting the dependencies between variables. A widely used type of composition is the nonlinear weighted sum $f(X)=K\left[\sum_i w_i g_i(X)\right]$, where K is some predefined function, such as the hyperbolic tangent (widely used for climate data).

Learning: Given a specific *task* to solve, and a *class* of functions F, learning means using a set of *observations*, in order to find $f^*\in F$, which solves the task in an *optimal sense*. This entails defining a cost function $C : F\rightarrow\Re$ such that, for the optimal solution f^*, $C(f^*)\leq C(f)\ \forall f\in F$, i.e., no solution has a cost less than the cost of the optimal solution. The cost function is an important concept in learning, as it is a measure of how far away we are from an optimal solution to the problem that we want to solve. Learning algorithms search through the solution space in order to find a function that has the smallest possible cost. For applications where the solution is dependent on some data, the cost must necessarily be a function of the observations, otherwise we would not be modelling anything related to the data. It is frequently defined as a statistic to which only approximations can be made. In this work we consider the problem of finding the model f which minimises the error function, for data pairs $(x\in X,\ y\in Y)$ drawn from some distribution D.

Pratically, the cost is minimised over a sample of the data rather than the true data distribution. Despite the three major learning paradigms, in supervised learning, we are given a set of ($x \in X$, $y \in Y$) pairs and the aim is to find a function f in the allowed class of functions that matches. In other words, we wish to *infer* the mapping implied by the data; the cost function is related to the mismatch between our mapping and the data and it implicitly contains prior knowledge about the problem domain. A commonly used cost is the mean-squared error which tries to minimise the average error between the network's output, $f(x)$, and the target value y over all pairs. When one tries to minimise this cost using gradient descent (done by simply taking the derivative of the cost function with respect to the network parameters and then changing those parameters in a gradient-related direction) for the class of neural networks called Multilayer Perceptrons, one obtains the well-known backpropagation algorithm for training neural networks. Tasks that fall within the paradigm of supervised learning are pattern recognition (also known as classification) and regression (also known as function approximation). The supervised learning paradigm is widely applicable to sequential data. The algorithms are mathematical techniques for minimising the discrepancy between a parameterised function and a set of pairs of inputs and "correct" outputs, where the overall function is partitioned into layers of vector functions.

Back Propagation: It is the best-known training algorithm for multi-layer neural networks. It defines rules of propagating the network error back from network output to network input units and adjusting network weights along with this back propagation. It requires lower memory resources than most learning algorithms and usually gets an acceptable result, although it can be too slow to reach the error minimum and sometimes does not find the best solution.

Quick Propagation: It is a heuristic modification of the back propagation algorithm. This training algorithm treats the weights as if they were quasi-independent and attempts to use a simple quadratic model to approximate the error surface. In spite the fact that the algorithm hasn't got theoretical foundation, it's proved to be much faster than standard backpropagation for many problems. However, sometimes the quick propagation algorithm may be unstable and inclined to stuck in local minima.

Training a neural network model essentially means selecting one model from the set of allowed models, i.e., in a Bayesian framework, determining a distribution over the set of allowed models that minimises the cost criterion. Evolutionary methods, simulated annealing, and expectation-maximisation and non-parametric methods are among other commonly used methods for training neural networks.

The stochastic ANN approach using empirical Bayesian updating seems to be an important tool for the propagation of the related weather information to provide practical geostatistics solution of uncertainties associated with the interpolation and capturing the spatiotemporal structure of the data. The basic idea is to import the entire posterior distribution from other locations allowing prediction of unsampled weather parameters using spatial related sampled information. The temporal dependence of model parameters is evaluated in a Bayesian framework. A model is used to predict the process of interest Y at the time t. This multivariate procedure uses the available related weather data sets and climate proxies (monitoring and assembling network sites). The ANN methodology is applied to climate time series (regional precipitation records, using climate proxies). In particular, cross-correlation technique is applied to examine coherence and phase relationships between various climate time series on interannual scale. A model is used to predict the process Y at the time t and location s.

Lemma 1: The prior information at time t can be modelled by a temporal prior function given by:

$$\Theta(Y(t)) = f(Y(t)|X(t_1), ..., X(t_k)).$$

Lemma 2: The model of temporal dependence allows an empirical Bayesian updating of any prior $\Theta(Y(t))$ by neighbouring related data $s_1, ..., s_n$.

The basic idea is to interpret the prior distribution $\Theta(y \in Y(t))$ as realisations of the corresponding temporal random function $\Theta(Y(t))$. The spatiotemporal dependence can be explored by examining the distribution of nearest-neighbour distances.

4 Experimental dataset and missing data estimation

The ANN technique is illustrated by means of some real case studies of precipitation through the state of the Rio Grande do Sul (RS) in Brazil (Fig. 1), taking into account the monthly total rainfall series recorded in the period 1961–2005. The meteorological stations 83967 (Porto Alegre), 83907 (São Luiz Gonzaga), 83914 (Passo Fundo) and 83980 (Bagé) are the time series that presented smaller number of missing values – that's because they have been chosen as "control" for the local information, to verify the behavior and the performance of the trained network. In a similar way, it was observed that the meteorological station 83919 (Bom Jesus) presented a total of 95 continuous missing values in the beginning of the series and the station 83995 (Rio Grande) include a total of 49 intermittent missing values – these are our "target" stations!

In any spatial interpolation scheme the selection and quantity of surrounding stations are critically important to the results of the interpolations. Problems arise when using climatological data because of missing values and the varying availability of stations through time. In order to determine which stations are to be used, surrounding stations are preselected based on their relationship with the target station. The closest stations are identified for each target station and are ranked by the value of the correlation coefficient between

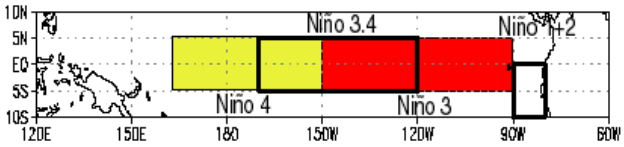

Fig. 2. The Niños Region: Niño1+2: 0–10 OS, 90–80 OW, Niño3: 5 ON–5 OS, 150–90 OW, Niño4: 5 ON–5 OS, 160 E–150 OW, Niño3+4: 5 ON–5 OS, 170–120 OW.

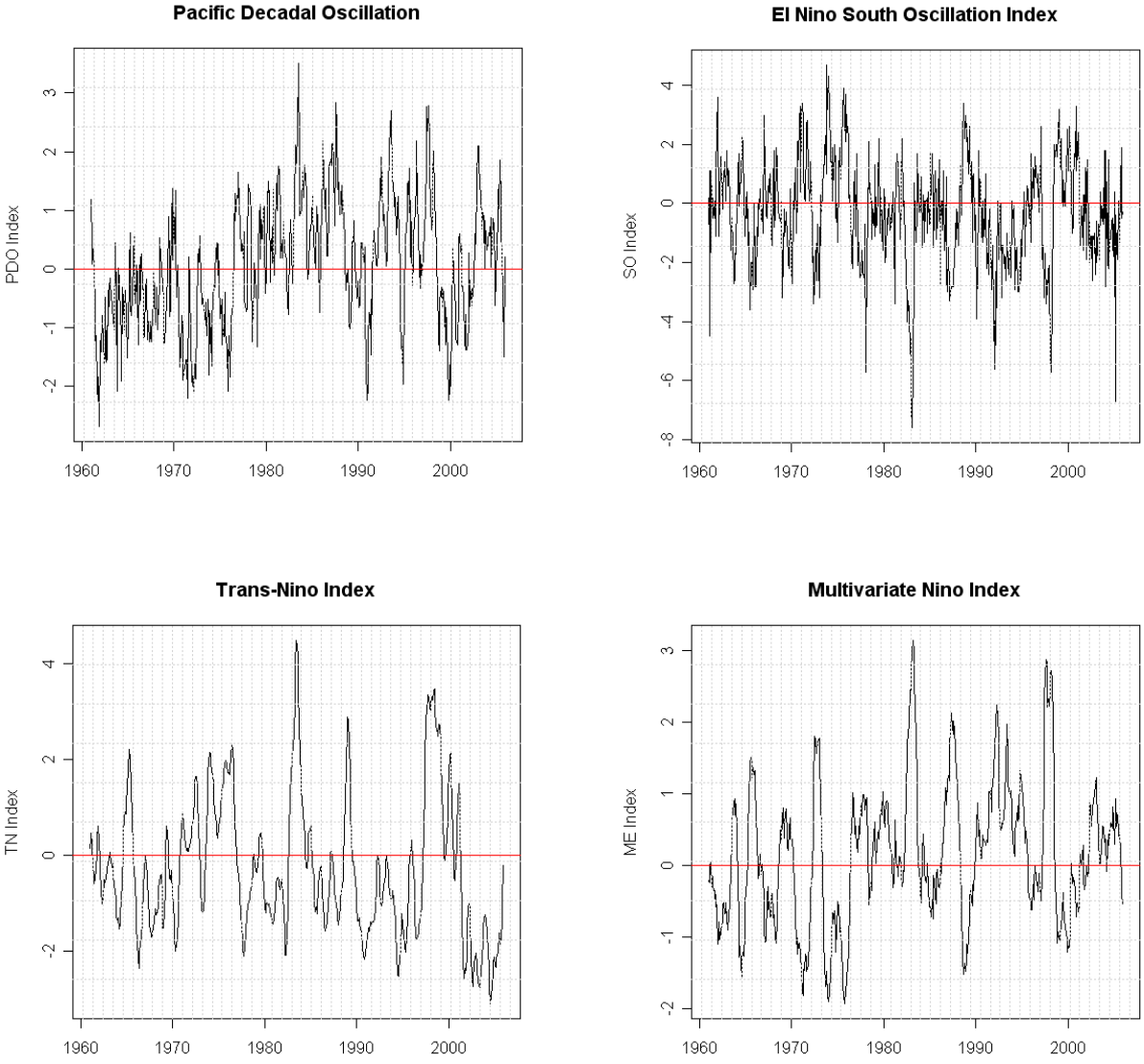

Fig. 3. The time series graphic representation of the monthly "climate proxy: 4 indicators employed in this manuscript.

the candidate station and its neighbours. The ANN estimation technique based on spatiotemporal objective analysis scheme is used to estimate monthly values, with the "best" estimate chosen as a missing value replacement for the development of regional monthly total precipitation time series over Brazil.

One determines the embedding dimension (number of past observations) of the time series attractor (delay time that determine how data are processed) and uses this measure to define the network's architecture. Physically, the attractor is the object to which the time series in a phase space (space in which each point describes the state of a dynamical system

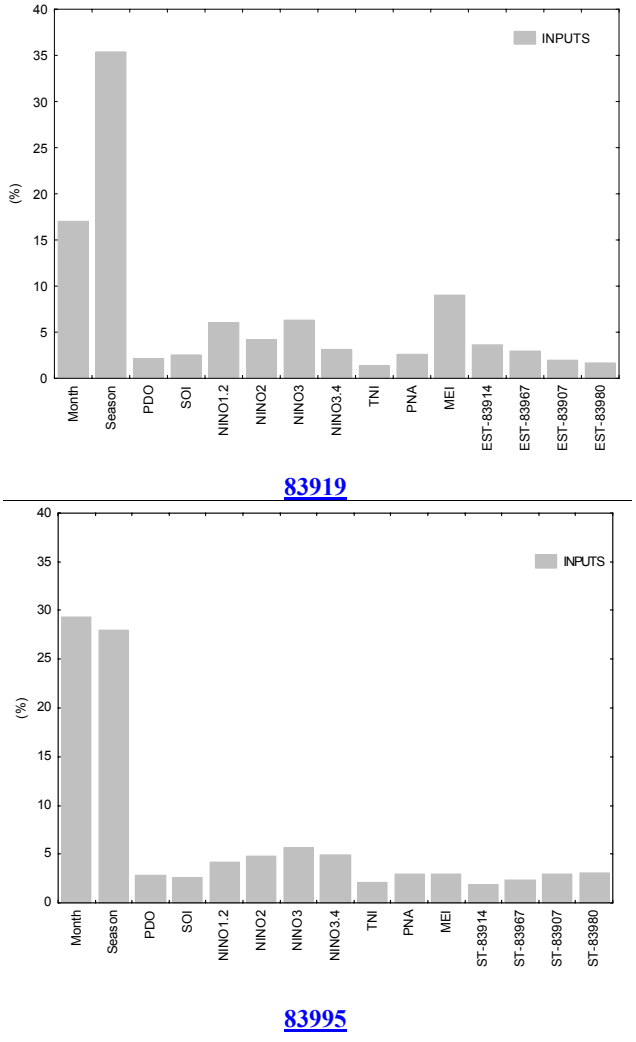

Fig. 4. Sensibility analysis of the input variables in the precipitation times series reconstruction, via ANN, of the meteorological stations 83919 and 83995.

as a function of the non-constant parameters of the system) is attracted to. Some meteorological attributes can be accurately predicted by the spatiotemporal ANN model architecture: designing, training, validation and testing. The best generalisation of new data is obtained when the mapping represents the systematic aspects of the data, rather capturing the specific details of the particular training set.

The replacement of missing monthly values for total precipitation includes the use of nearby simultaneous values to calculate an estimated value at the target station over the period of time for which adequate data are available. The efficiency, or accuracy, of the estimates over a long period of time provides the information used to assess the quality of estimated monthly values. Estimated monthly values are used in "*lieu*" of missing values as a mean of making a particular station serially complete. There are numerous spatial interpolation methods available for point estimation with irregularly spaced data. Typically, the choice of methodology is dependent on several factors: the meteorological variable under consideration, the geographical area, the spatial distribution of surrounding observations, and the day–month–season for which the target station will be estimated.

5 Large-scale teleconnections

Climate proxies are sources of climate information and variability from natural archives such as historical records, which can be used to estimate climate conditions. The proxy indicators typically must be calibrated to yield a quantitative reconstruction of past climate. In effect, a proxy variable is something that is probably not in itself of any direct interest, but from which a variable of interest can be obtained. In this work we use the information of particular sea surface temperature (Fig. 2) and some climate proxies (Fig. 3) of ocean characteristic to "create" rain-gauge based precipitation records. These indicators are considered in the first (input) layer of the ANN, where it is necessary to carefully calibrate the proxy against the variable of interest, in this case local monthly precipitation.

The Southern Oscillation Index (SOI) is based on the standardised pressure difference between Tahiti and Darwin. The El Niño Southern Oscillation (ENSO) phenomenon is the major cause of year-to-year variations in climate over the globe. The Pacific Decadal Oscillation (PDO) is a leading index associated to the ENSO phenomenon by taking into account the monthly Sea Surface Temperature (SST) anomalies in the North Pacific Ocean. In effect, to characterise the nature of the ENSO, SST anomalies in different regions of the Pacific is used.

The Trans-Niño Index (TNI), which is given by the difference in standardised (1950–1979) anomalies of SST between Niño1+2 and Niño4 regions, is used as an optimal description of the character and evolution of El Niño or La Niña. The Pacific Decadal Oscillation (PDO) is a leading index associated with the Sea Surface Temperature (SST) anomalies in the North Pacific Ocean. In effect, to characterise the nature of the ENSO, SST anomalies in different regions of the Pacific is used.

The Multivariate ENSO Index (MEI) is calculated based on the six main observed variables over the tropical Pacific. These six variables are: sea-level pressure (P), zonal (U) and meridional (V) components of the surface wind, sea surface temperature (S), surface air temperature (A), and total cloudiness fraction of the sky (C). The MEI is computed separately for each of twelve sliding bi-monthly seasons (Dec/Jan, Jan/Feb,..., Nov/Dec). The MEI is calculated as the first unrotated Principal Component (PC) of all six observed fields combined.

The PNA teleconnection index, a measure of the strength and phase of the Pacific/North American teleconnection

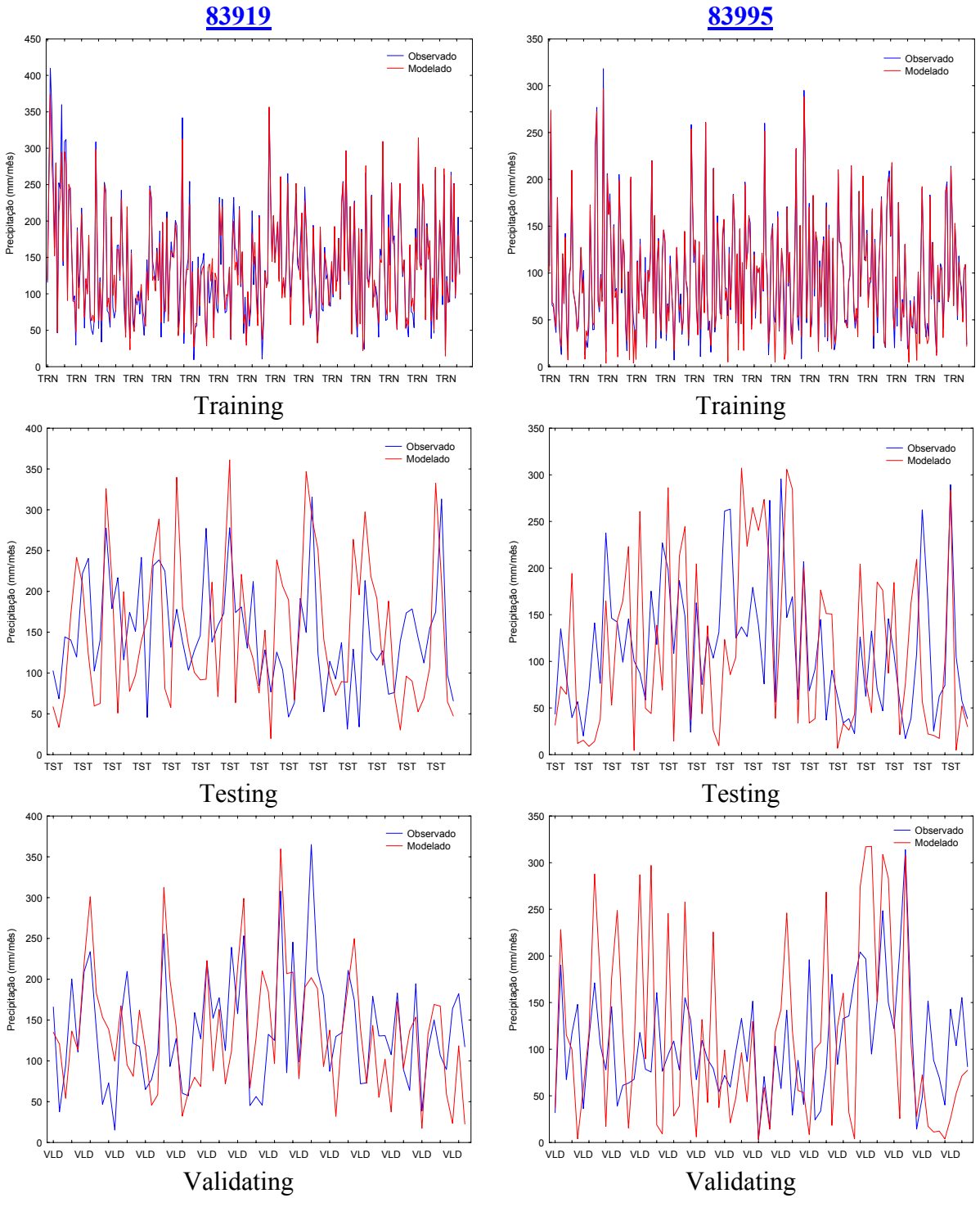

Fig. 5. 83919 – ANN Architecture: [15-76-1], "Bootstrap" resampling rate: $|W|$=0.8% (percentage of the cross-validation) and 83995 – ANN Architecture: [15-39-1], $|W|$=1.4%. Hyperbolic tangent, Quick Propagation algorithm. Learning rate was set to 1%; momentum factor rate: 1.75. 2000 iterations and 2 retraining. 68% of the data were used for training (TRN) the network, 16% were randomly selected to be used as test (TST) pattern and 16% were used for validation (VLD).

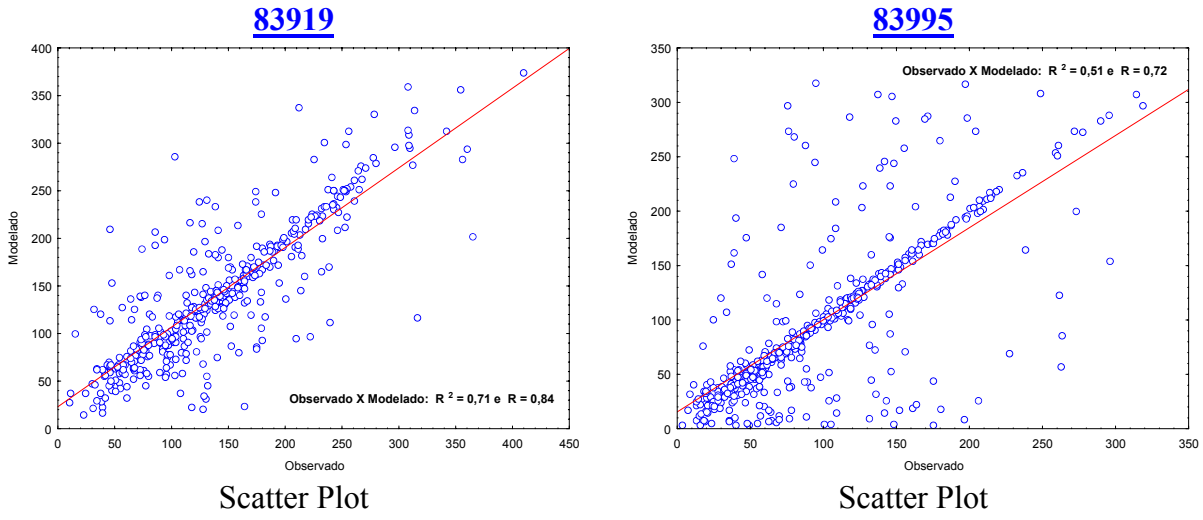

Fig. 6. Mapping learning with correlation coefficient of about 0.84 for 83919 and 0.72 for 83995.

pattern, is used to examine changes in the midtropospheric flow over North America on decadal, interannual, and intra-annual time scales. The index corroborates previous findings that a major change in the midtropospheric circulation took place over North America during the late 1950s. The time series of index values also demonstrates the existence of a previously unknown quasi periodicity in the configuration of midtropospheric heights over the North American sector (Leathers and Palecki, 1991). A different formulation of the PNA index, namely as the second principal component of Northern Hemisphere extratropical sea-level pressure anomalies, was proposed in the 1990s. The Northern Annular Mode (NAM) is the first Empirical Orthogonal Function (EOF), which explains the maximum variability of the process (EOF1) and the PNA is the EOF2

In practice, one splits the available data in training, testing and validating (evaluation) sets. For each proxy vector the corresponding rain-gauge values on the ground were also known in neighbour related or control stations. The rain-gauge values were used as the table of truth in order to decide whether the rain prediction using the proxies' measurements was successful.

6 Results and conclusions

This research work summarizes a procedure used to create serially complete monthly precipitation datasets (1961–2005) for the state of Rio Grande do Sul - Brazil. Determining target and estimator stations by scanning the quality of individual station records, reconciling metadata (including observation times and station locations), and categorising observations proved to be time consuming but necessary. Estimating the missing data values and cross validating the results proved to be relatively straightforward once preparatory work was accomplished. Our results show that the efficacy of the estimation procedure and thus the reliability of the estimated missing values are dependent on a number of factors.

In this study, different neural network architectures and learning rates were tested, aiming at establishing a network that resulted in the best possible reconstruction of missing rainfall data. A multiple hidden layer architecture was chosen. This kind of architecture has been adopted for solving problems with similar requirements. The parameters used for the training of the network were collected at each control station and climate proxies. It is well known that depending on the meteorological parameter under study, the selection and quantity of surrounding stations are critically important to the results of the interpolations. We feel that the pre-selection of surrounding stations, based on their relationship with the station to be estimated, is an integral first step.

Through inputs importance (sensibility analysis) – Fig. 4, it was verified that the periodicity variables (month and season) were the most important for the network during the training phase in both the stations (greater than 50%). Although should be emphasised that for the station 83919 the proxy MEI presented an importance of 10% while for the station 83995 the variable NINO3 was the most important proxy (6%). In Fig. 5 and 6 were verified for the station 83919 a correlation of 0,84 and R^2 of 0,72 with the selected neural network architecture [15(29)-76-1] – input: 15 attributes (13 continuos variables = 4 rain-gauge stations, 5 climate proxies and 4 sea surface temperature from Niño regions; 2 categorical variables: months and seasons) linear neurons (input: 29 (13+12+4) attributes if we consider that each categorical variable can be represented by each intrinsic classification) – hidden: 76 neurons (hyperbolic tangent with Gaussian complement) – output: 1 linear neuron (with logistic control) –

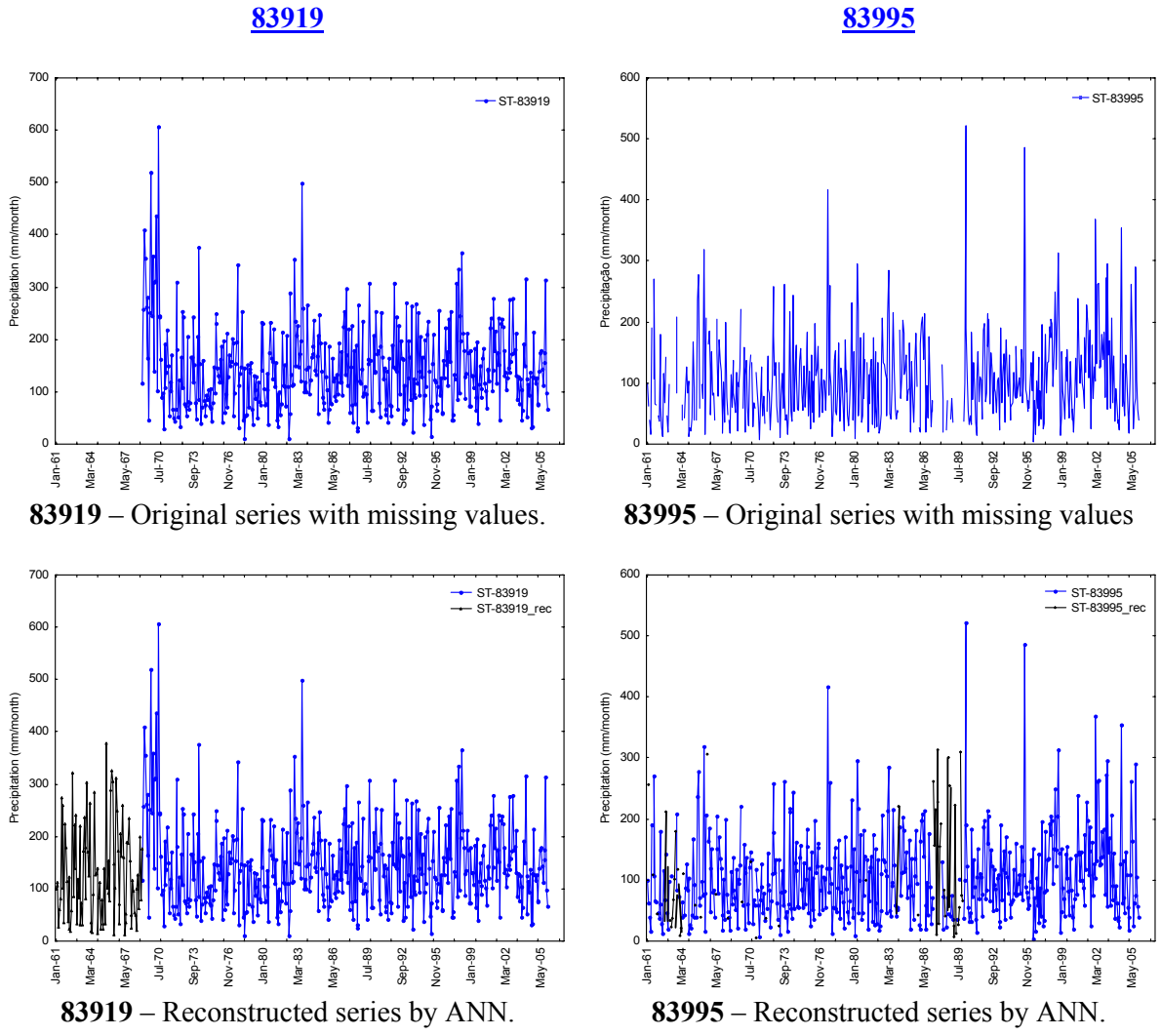

Fig. 7. Time series of monthly precipitation: Original versus Reconstructed.

while for the station 83995 presented a correlation of 0.72 and R^2 of 0.51 with [15-39-1]. It was demonstrated that the network presented a optimum answer during the training phase, however, during the validation and test phase the net tends to overestimate the observed values, mostly, at the station 83995. Subsequent to the net calibration and training was accomplished the series reconstruction of monthly precipitation to stations 83919 (81 missing values at the beginning of the series) and 83995 (49 missing values intermittent).

On the other hand, should be carried in consideration that the trained network just results from some variables "proxies" and rain-gauge data from 4 control stations of input (without missing values or maximum of 10 missing values inside the series). Maybe, new variables inclusion (information locals in situ) in the training phase will be able to contribute for a more robust form for the results generated by the model. The correlation coefficient obtained for the training data set was about 70%. The verification of the network was done by using unknown data for the target station. This was made for months, whose data were excluded from the training set. The correlation coefficient for the unknown case was about 80%.

The conclusions highlight the use of climate proxies response as potential weather predictor. The use of ANN has been recognised recently as a promising way of making predictions on time series, detecting irregular behaviour. As expected, this robust reconstruction method has good performance; since more information is introduced in the decision-making system (cf., Fig. 7). The technique introduced in the present study aims primarily at producing a spatiotemporal series of monthly rainfall at an observation site with a limited set of data. The findings presented in this study show that the series reconstruction using artificial neural networks for the

intended purpose is adequately acceptable. The prediction error was confined to less than 5%, which is considered by the meteorologist satisfactory.

Acknowledgements. Grateful thanks to Alex Grechanowski for kindly providing us the software NeuroIntelligence – Neural Network Software for Professionals. Grateful thanks are also due to the referees and the journal editors. Their comments were very useful in improving this article.

Edited by: S. C. Michaelides and E. Amitai
Reviewed by: anonymous referees

References

Abdelaal, R. E. and Elhadidy, M. A.: Modelling and forecasting the daily maximum temperature using adductive machine learning, Wea. Forecast., 10, 310–325, 1995.

Alexandersson, H.: Homogeneity testing, multiple breaks and trends', Proceedings of the 6th International Meeting on Statistical Climatology, Galway, 439–441, 1995.

Alexandersson, H. and Moberg, A.: Homogenization of Swedish temperature data. Part I: A homogeneity test for linear trends, Int. J. Climatol., 17, 25–34, 1997.

Alexandersson, H.: A homogeneity test applied to precipitation data, J. Climatol., 6, 661–675, 1986.

Andreoli, R. V. and Kayano, M. T.: ENSO-related rainfall anomalies in South America and associated circulation features during warm and cold Pacific decadal oscillation regimes, Int. J. Climatol., 25, 2017–2030, 2005.

Bishop, C. M.: Neural Networks for Pattern Recognition, Oxford, Oxford University Press, 1995.

Briggs, W. M. and Wilks, D. S.: Estimating monthly and seasonal distributions of temperature and precipitation using the new CPC long-range forecasts, J. Climate, 9, 818–839, 1996.

Buishand, T.: Some methods for testing the homogeneity of rainfall records, J. Hydrol.., 58, 11–27, 1982.

Chow, G. C.: Tests of Equality Between Sets of Coefficients in Two Linear Regressions, Econometrica, 1960.

Craddock, J. M.: Methods of comparing annual rainfall records for climatic purposes, Weather, 34, 332–346, 1979.

DeGaetano, A. T., Eggleston, K. L., and Knapp, W. W.: A method to produce serially complete daily maximum and minimum temperature data for the Northeast, NRCC Research Publication RR 93-2, 9 pp., 1993.

Eischeid, J. K., Plantico, M., and Lott, N.: Creating a Serially Complete National Daily Time Series of Temperature and Precipitation for the Western United States, J. Appl. Meteorol., 39, 1580–1591, 2000.

Grimm, A. M., Barros, V. R., and Doyle, M. E.: Climate variability in Southern South America associated with El Niño and La Niña events, J. Climate, 13, 35–58, 2000.

Huth, R. and Nemesova, I.: Estimation of missing daily temperatures: Can a weather categorization improve its accuracy?, J. Climate, 8, 1901–1916, 1995

Kalogirou, S., Neocleous, C., Michaelides, S., and Schizas, C.: A Time Series Reconstruction of Precipitation Records Using Artificial Neural Networks, Proceedings of the EUFIT'97 Conference, Aachen, Germany, 3, 2409–2413, 1997.

Kaplan, A., Kushnir, Y., and Cane, M. A.: Reduced Space Optimal Interpolation of Historical Marine Sea Level Pressure: 1854–1992, J. Climate, 13, 2987–3002, 2000.

Knippertz, P., Christoph, M., and Speth, P.: Long-term precipitation variability in Morocco and the link to the large-scale circulation in recent and future climates, Meteorol. Atmos. Phys., 83, 67–88, 2003.

Kousky, V. E., Kayano, M. T., and Cavalcanti, I. F. A.: A review of the Southern Oscillation: oceanic- atmospheric circulation changes and related rainfall anomalies, Tellus, 36A, 490–504, 1984.

Kyriakidis, P. C. and Journel, A. G.: Geostatistical space-time models: a review, Math. Geol., 31(6), 651–684, 1999.

Leathers, D. J. and Palecki, M. A.: The Pacific/North American Teleconnection Pattern and United States Climate. Part II: Temporal Characteristics and Index Specification, J. Climate, 5, 7, 707–716, 1991.

Michelangeli, P., Vautard, R., and Legras, B.: Weather regimes: recurrence and quasi-stationarity, J. Atmos. Sci., 52, 1237–1256, 1995.

Michaelides, S. C., Neocleous, C. C., and Schizas, C. N.: Artificial neural networks and multiple linear regression in estimating missing rainfall data, Proceedings of the DSP95 International Conference on Digital Signal Processing, Limassol, Cyprus, 668–673, 1995.

Murphy, J.: An evaluation of statistical and dynamical techniques for downscaling local climate, Int. J. Climatol., 12, 2256–2284, 1999.

Pettitt, A. N.: A Non-Parametric Approach to the Change-Point Problem, Appl. Stat., 28(2), 126–135, 1979.

Schizas, C. N., Pattichis, C. S., and Michaelides, S. C.: Forecasting minimum temperature with short time-length data using artificial neural networks, Neural Network World, 94(2), 219–230, 1994.

Shumway, R. H. and Stoffer, D. S.: Time series analysis and its applications, New York, Springer-Verlag, 2000.

von Neumann, J.: Distribution of the ratio of the mean square successive difference to the variance, Ann. Math. Stat., 12, 367–395, 1941.

Wilks, D. S.: Statistical significance of long-range optimal climate normal temperature and precipitation forecasts, J. Climate, 9, 827–839, 1996.

Zorita, E., Hughes, J. P., Lettemaier, D. P., and von Storch, H.: Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation, J. Climate, 8, 1023–1042, 1995.